**PREDICTING OPTIMAL LOCATION FOR A RESTAURANT**

**1. INTRODUCTION TO THE BUSINESS PROBLEM**

**1.1 BACKGROUND**

Alberta is a province in Canada. It has a population of about 4 million and is termed as Canada's fourth most populous province. It includes 19 cities like St. Albert, Edmonton, Lethbridge, Calgary, Red Deer, Brooks, etc. Calgary is the largest city of Alberta. It has population exceeding 1 million and is the home to Canada's second-highest number of corporate head offices among the country's 800 largest corporations.

**1.2 PROBLEM**

This project will try to find an optimal location for opening a multi-cuisine restaurant in Calgary (as it is the largest city and also the corporate hub). Optimal location implies that there are fewer numbers of restaurants of the same type. Also, it will be beneficial if the location is a hub of corporate offices or colleges.

**1.3 INTEREST**

This report will specifically be targeted to the stakeholders who are looking for opportunities in opening a restaurant in Calgary.

**2. DATA DESCRIPTION**

Based on the above problem, we require the number, names and the type of restaurants present in Calgary.

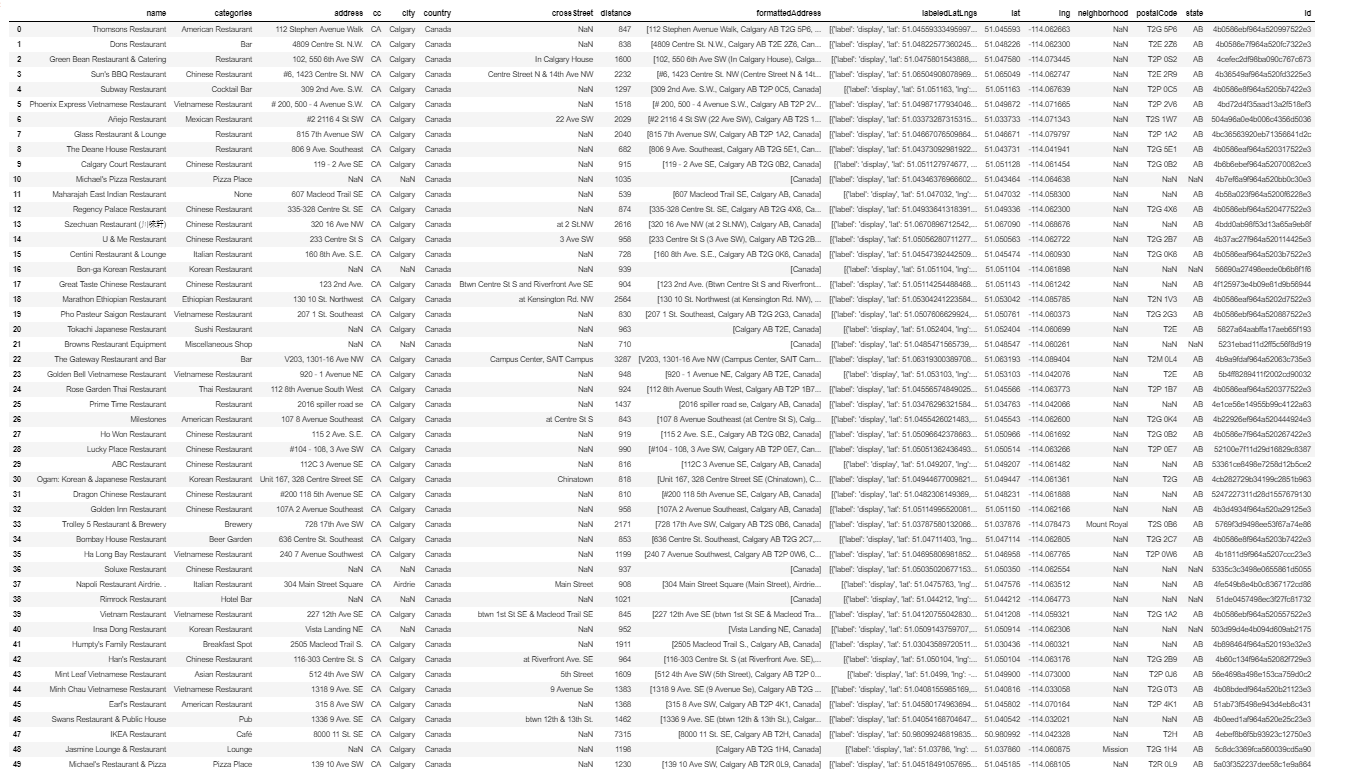
Firstly, using the Foursquare API, the latitude and longitude of Downtown (Centre of Calgary) was extracted. Then, used those coordinates to find other present restaurants within a certain radius. The returned json file was then turned into a data frame.

**2.1 DATA CLEANING**

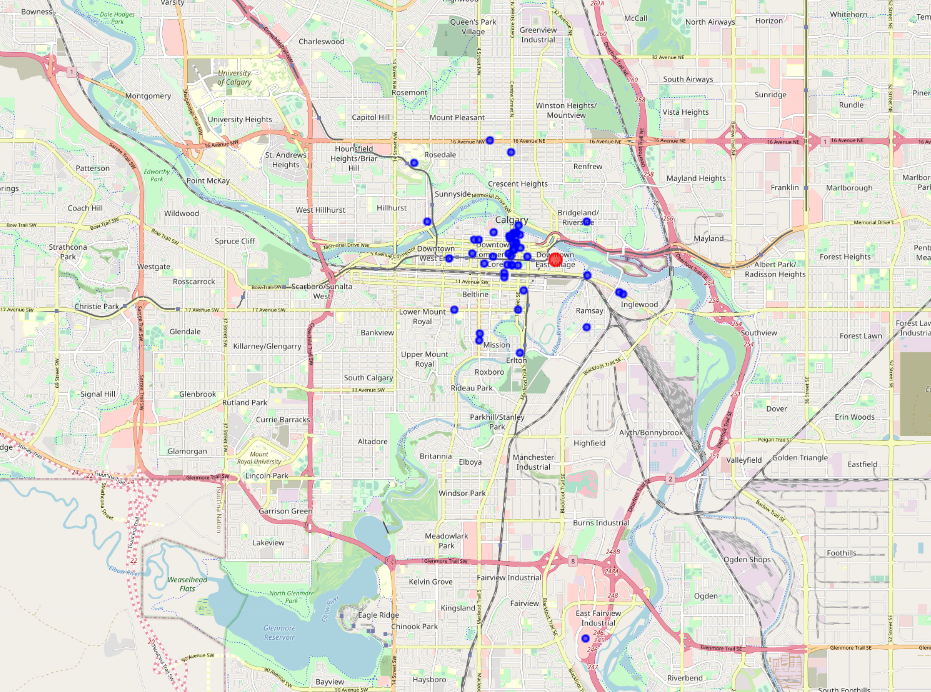
The data frame looked like:



It was then transformed according to the needs. The redundant columns like Perk, referral id, and venue page id were removed. In the rest of the columns, location. was removed and the category type was extracted from a long tuple of values in the category column. The final data frame looked like:



We then visualized these restaurants on the map of Calgary. Red represents Downtown and blue shows the restaurants surrounding it.



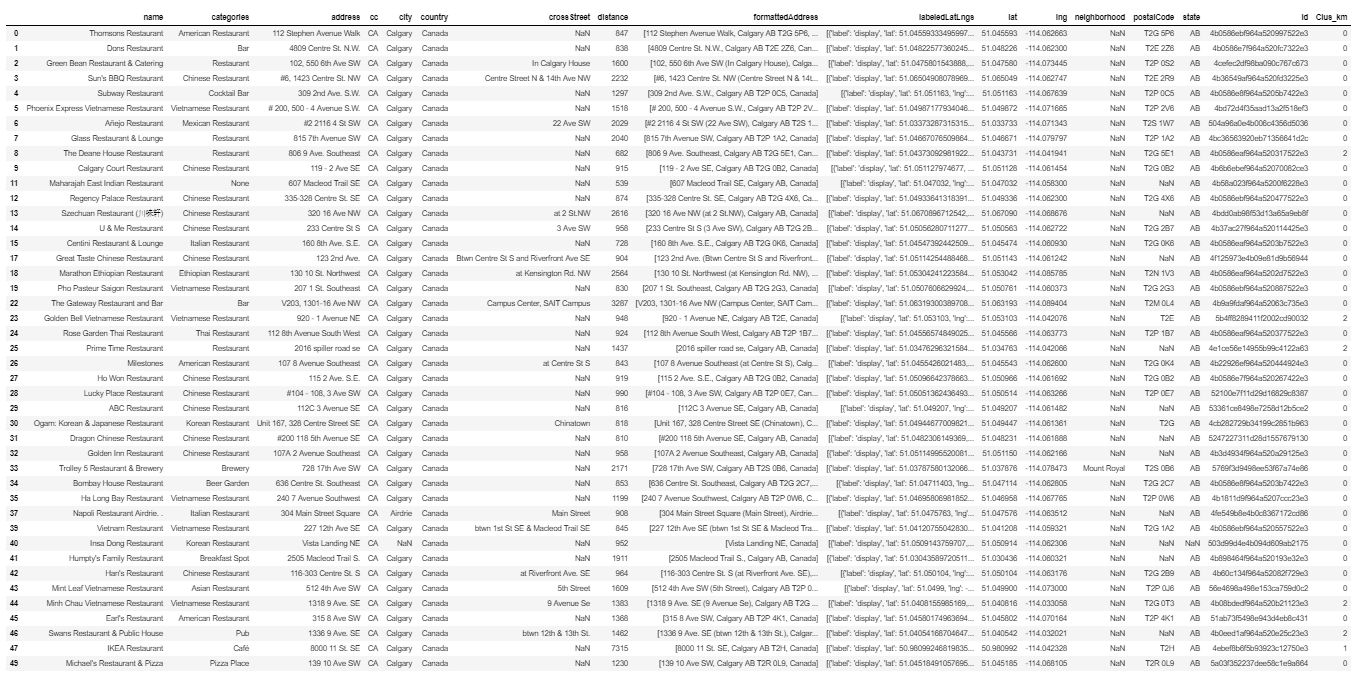
**3. DATA MODELLING**

To find the optimal location for a restaurant, the effort was put into clustering the present restaurants based on their distance from the Downtown. Restaurants lying very near to downtown were clustered into one, restaurants a little far away were clustered together and so on.

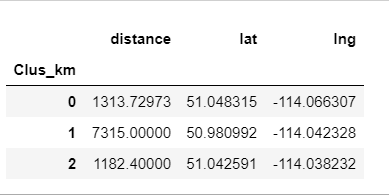
**3.1 ML ALGORITHM USED**

K-means algorithm of machine learning which is used for clustering was used in this project. This algorithm divided the places into k (user defined) groups.

For applying k-means, the first step was to normalize the data which was accomplished using Standard scaler library. After normalizing, the restaurants were divided into 3 clusters using k-means and labels were generated. These labels were then assigned to their respective restaurants and appended in the table.

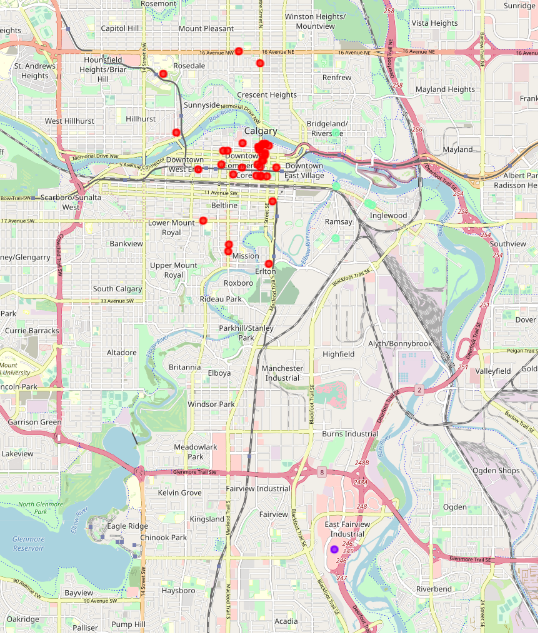


The data frame was then grouped on the basis of cluster labels giving the mean distance of all the restaurants in a cluster from downtown.



**3.2 DATA VISUALIZATION**

After the clustering has been done, visualization of the clusters on the map of Calgary was done. We can clearly see that the restaurants were divided into three clusters (shown in red, green and blue).



**4. RESULTS**

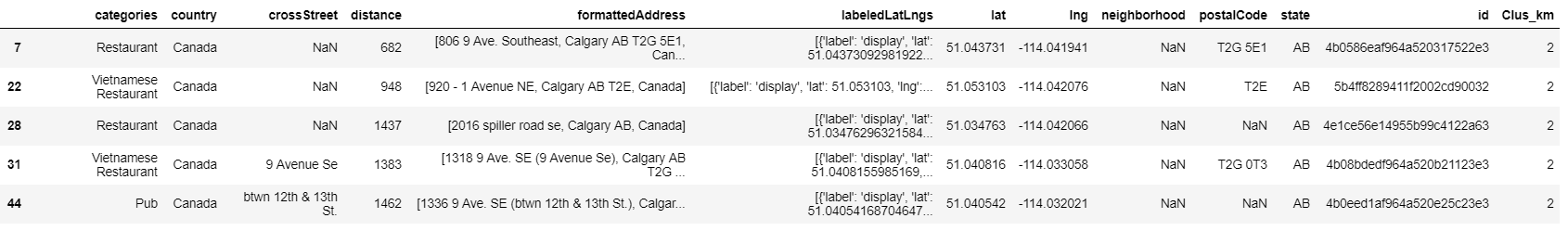
This analysis shows that there are 37 restaurants in cluster 0, 1 in cluster 1 and 5 in cluster 2.

Cluster 0:

Cluster 1:



Cluster 2:



**5. CONCLUSION**

After all the modelling and analysis, I came to a conclusion that there are lesser number of restaurants in the south of Calgary. So, if anyone wishes to open a multi cuisine restaurant in Calgary, may open in areas like Kingsland, Meadowlark Park, Chinook, and Fairview. These areas are also surrounded by industrial areas which will help in growing the business.